A French Interaction Grammar

Guy Perrier

A French Interaction Grammar

Guy Perrier - LORIA, université Nancy2 (France)

Interaction Grammars (IGs) are a grammatical formalism which uses two fundamental concepts : *underspecification* and *polarity* [Per00, Per04]. These concepts apply to both syntax and semantics of natural languages but here, we only consider the syntactic level. In this context, underspecification essentially means underspecification of syntactic trees, and it is expressed using the notion of *tree description*. A tree description is a flexible and compact way of representing a family of syntactic trees sharing some properties. By decorating the descriptions with polarized features, we can express the valences of the syntactic trees : a positive feature represents an available resource whereas a negative feature represents an expected resource.

Syntactic composition consists of superposing tree descriptions while respecting polarities: a negative feature must encounter a dual positive feature to be neutralized and vice versa. Parsing a sentence can be compared to an electrostatic process : since IGs are lexicalized, a lexicon provides a polarized tree description for every word of the sentence and then, we have to superpose the selected descriptions to build a completely specified tree, where all polarities are neutralized.

To summarize, IGs combine the flexibility of Unification Grammars with the saturation control of Categorial Grammars.

Polarities allow us to develop original methods of parsing, which are implemented in the parser LEOPAR[1] : the filtering of lexical selections is performed by automata which take only polarities into account from the selected tree descriptions : they ignore structural information; then, the parsing process itself, also based on the neutralization of opposite features, takes structural constraints into account with different possible strategies : top-down, bottom-up, incremental strategies.

To test LEOPAR, we are currently developing a large scale French interaction grammar with two principles:

- The grammar is completely lexicalized, so that every elementary tree description of the grammar is anchored by a word of the language. But the anchoring mechanism aims at the re-usability of lexical information. For this, we have built a syntactic lexicon which is independent of the IG formalism (this lexicon was used for TAG). Every entry of this lexicon associates an inflected word with a feature structure, describing the morphological and syntactic properties of the word in a way which is completely neutral with respect to any formalism.

  In the grammar, every tree description is equipped with an interface in the form of a feature structure, which gives the properties of all the words that are able to anchor this description. Then, anchoring is performed by unification between the entries of the lexicon and the interfaces of descriptions from the grammar. The advantage is that the same lexicon can be used for various grammatical formalisms.

[1] `www.loria.fr/equipes/calligramme/leopar`

- Concerning the design of the grammar, a fundamental idea, which underlies the concept of *meta-grammar*, is to dissociate the level at which the grammar is written from the level at which the grammar is used in NLP systems. In a way, we use the same approach as in the design of programming languages with the distinction between source programs and object programs. For this, we use XMG [2], a platform which allows the development of grammars in various formalisms (it is also used for TAG). Grammars are modular and possibly multidimensional; modules can be combined with two operators: conjunction and disjunction. Then, the terminal modules of the hierarchy are compiled into a grammar, which is directly usable in NLP systems.

The already built grammar includes 400 modules, among which 100 terminal modules; XMG compiles these 100 modules into 800 tree descriptions. The grammar was tested on the TSNLP corpus [LORP+96]. This corpus includes 1690 acceptable sentences and 1935 unacceptable sentences. These sentences were parsed with LEOPAR and the French grammar. The grammar correctly analyses 82% of the acceptable sentences and correctly identifies (i.e. fails to parse) 86% of the unacceptable sentences. For grammatical items, the failures are due to the fact that the grammar does not take into account spoken language, frozen expressions, adverbs and adjuncts in non standard positions, specific grammatical phenomena (causative constructions, superlatives, polite plural). The 14% of unacceptable sentences that are parsed successfully, are unacceptable for semantic or phonological reasons, whereas our grammar only covers French syntax.

In terms of expressiveness, the strong points of IG are: long distance dependencies (pied-piping, barriers to extraction ...), negation (for which the existence of pairs such as *ne ... aucun* with a relatively free position for *aucun* constitutes a challenge); coordination is also a phenonemon that can be modelled successfully in IG.

Difficult phenomena in French, such as subject inversion or agreement with the past participle in combination with the auxiliary *avoir*, are partially modelled in our grammars.

Frozen expressions, adverbs and adjuncts in non standard positions, parentheticals pose difficult problems, which are not solved in the current grammar.

References

[LORP+96] Sabine Lehmann, Stephan Oepen, Sylvie Regnier-Prost, Klaus Netter, Veronika Lux, Judith Klein, Kirsten Falkedal, Frederik Fouvry, Dominique Estival, Eva Dauphin, Hervé Compagnion, Judith Baur, Lorna Balkan, and Doug Arnold. TSNLP — Test Suites for Natural Language Processing. In *Proceedings of COLING 1996,Kopenhagen*, 1996.

[Per00] G. Perrier. Interaction grammars. In *COLING 2000, Saarbrücken*, pages 600–606, 2000.

[Per04] G. Perrier. La sémantique dans les grammaires d'interaction. *Traitement Automatique des Langues*, 45(3):123–144, 2004.

[2] `sourcesup.cru.fr/projects/xmg`